

**Application of Bayesian model averaging to measurements of the primordial power spectrum**

David Parkinson and Andrew R. Liddle

*Astronomy Centre, University of Sussex, Brighton BN1 9QH, United Kingdom*

(Received 8 September 2010; published 24 November 2010)

Cosmological parameter uncertainties are often stated assuming a particular model, neglecting the model uncertainty, even when Bayesian model selection is unable to identify a conclusive best model. Bayesian model averaging is a method for assessing parameter uncertainties in situations where there is also uncertainty in the underlying model. We apply model averaging to the estimation of the parameters associated with the primordial power spectra of curvature and tensor perturbations. We use COSMONEST and MULTINEST to compute the model evidences and posteriors, using cosmic microwave data from WMAP, ACBAR, BOOMERanG, and CBI, plus large-scale structure data from the SDSS DR7. We find that the model-averaged 95% credible interval for the spectral index using all of the data is  $0.940 < n_s < 1.000$ , where  $n_s$  is specified at a pivot scale  $0.015 \text{ Mpc}^{-1}$ . For the tensors model averaging can tighten the credible upper limit, depending on prior assumptions.

DOI: 10.1103/PhysRevD.82.103533

PACS numbers: 98.80.-k

**I. INTRODUCTION**

Measurements of the cosmic microwave background (CMB) by the Wilkinson Microwave Anisotropy Probe (WMAP) have, over the last few years [1–3], produced ever more refined constraints on the cosmological parameters, particularly those relating to the spectrum of primordial perturbations. Typically, when constraints are quoted this is done under the assumption that a particular underlying model is the correct one, but there remains some uncertainty as to which is the most appropriate model to fit to the data. One approach to this is model selection, which asks the data to rank the models under consideration, but current applications [4–9] indicate that several models are still allowed.

Such uncertainty in the correct choice of model can be handled by the technique of *Bayesian model averaging* [10], which allows one to assess parameter uncertainties in the presence of model uncertainty. The individual posteriors from different models will contribute to the model-averaged posterior, weighted by their model likelihood. Our aim in this paper is to apply this technique to measurements of the primordial spectrum. In Sec. II we review the Bayesian ideas we are applying to measurements of the primordial power spectrum. In Sec. III we carry out the analysis.

**II. MULTI-MODEL BAYESIAN STATISTICS**

Bayes' theorem states the relationship between models ( $\mathcal{M}$ ), parameters ( $\bar{\theta}$ ), and data ( $D$ )

$$P(\bar{\theta}|D, \mathcal{M}) = \frac{P(D|\bar{\theta}, \mathcal{M})P(\bar{\theta}|\mathcal{M})}{P(D|\mathcal{M})}, \quad (1)$$

where  $P(\bar{\theta}|\mathcal{M})$  is the prior probability distribution of the parameters (assuming some model),  $P(D|\bar{\theta}, \mathcal{M})$  is the likelihood, and  $P(\bar{\theta}|D, \mathcal{M})$  is the posterior probability

distribution of the parameters. The prior is updated to the posterior by the likelihood. The term  $P(D|\mathcal{M})$  represents the model likelihood, and is called the *evidence*. In the case of single-model inference (where only a single model or set of parameters is considered), the evidence is simply a normalizing constant, set to satisfy the condition that the posterior distribution sums to unity.

However, in most interesting cases in cosmology, the correct model is not known, and the evidence takes different values for different models. We can use Bayes' theorem again, at one level above, to calculate the posterior odds between different models and perform *model selection*,

$$\frac{P(\mathcal{M}_1|D)}{P(\mathcal{M}_2|D)} = \frac{P(D|\mathcal{M}_1)}{P(D|\mathcal{M}_2)} \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \quad (2)$$

Here  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are the different models under consideration,  $P(\mathcal{M}_i)$  gives the prior probability of each model, and  $P(\mathcal{M}_i|D)$  is the model posterior probability. Thus, the model posterior probability is updated from the prior by the evidence. If the model priors are equal then the ratio of posteriors is simply the ratio of evidences. The ratio of evidences is commonly known as the Bayes factor  $B$  [11], and an interpretation scale was suggested by Jeffreys [12] (though some authors have started to use different language to qualify the different levels, e.g., Ref. [6]). Many papers have already been written about the use of the evidence for cosmological model selection [4–9, 13].

The logical procedure would be to first perform a model selection analysis to find the best model. Having done so, we would then perform parameter inference for the parameters of that single best model. However it is possible, even likely, that no model will have decisive evidence ( $\ln B > 5$ ) over all competing models. If we want to include this model uncertainty in the parameter posteriors, we could instead produce a model-averaged posterior distribution [10], where the individual posteriors from each

model are summed together, weighted by the model posterior values,

$$P(\bar{\theta}|D) = \frac{\sum_k P(\bar{\theta}|D, \mathcal{M}_k) P(\mathcal{M}_k|D)}{\sum_k P(\mathcal{M}_k|D)}. \quad (3)$$

This model-averaged posterior encodes the uncertainty as to the correct model.<sup>1</sup> This model-averaging procedure has been used before in cosmology [15] and astrophysics/geophysics [16].

### III. APPLICATION TO DATA

The primordial power spectrum of scalar perturbations is normally modeled through a modified power-law function of wave number  $k$ ,

$$\Delta_{\mathcal{R}}^2(k) = \Delta_{\mathcal{R}}^2(k_*) \left( \frac{k}{k_*} \right)^{(n_s-1)+(1/2)\ln(k/k_*)n_{\text{run}}}, \quad (4)$$

where the amplitude is defined at a pivot scale ( $k_*$ ),  $n_s$  is the spectral index (also known as the tilt), and  $n_{\text{run}}$  is the running of the spectral index. We also refer to  $\Delta_{\mathcal{R}}^2$  at the pivot scale as  $A_s$ . A maximally symmetric Harrison-Zel’dovich (HZ) [17] model has equal power on all scales, so the spectral index is unity and the running is zero. We discuss the choice of  $k_*$  below.

Inflation, currently our best model for the generation of the spectrum of Gaussian, adiabatic superhorizon perturbations, additionally predicts a spectrum of tensor perturbations, which are also modeled through a power law,

$$\Delta_h^2(k) = \Delta_h^2(k_*) \left( \frac{k}{k_*} \right)^{n_T}, \quad (5)$$

where  $n_T$  is the spectral index of the tensor perturbations (the tensor running is normally neglected). Single-field, slow-roll inflation predicts a consistency relation between the scalar and tensor amplitudes (measured at the same scale) in terms of the tensor spectral index,

$$\frac{\Delta_h^2(k_*)}{\Delta_{\mathcal{R}}^2(k_*)} \equiv r = -8n_T, \quad (6)$$

and we will enforce this throughout.

We considered five different models of the spectrum of primordial perturbations in this analysis:

- (I) A scale-invariant HZ spectrum of scalar perturbations with no tensor component ( $n_s = 1$ ,  $r = 0$ ).
- (II) A tilted model, where the spectral index is allowed to vary, still with no tensors.

<sup>1</sup>Though it may be that in the end the “true” model is not even one that we have considered at the time of the analysis. “Essentially, all models are wrong, but some are useful.” [14]

TABLE I. Prior ranges for the parameters in the different models. We considered only uniform priors in this analysis. The priors on power spectrum parameters are set at  $k_0 = 0.05 \text{ Mpc}^{-1}$ , the default for COSMOMC, but are so wide compared to the posteriors that the subsequent translation to the pivot scale is unaffected.

Models	Parameter	Min	Max
All	$\Omega_b h^2$	0.018	0.032
	$\Omega_c h^2$	0.04	0.16
	$\theta$	0.98	1.1
	$\tau$	0.01	0.3
	$\ln[10^{10} A_s]$	2.6	4.2
	$A_{\text{SZ}}$	0	2
tilt, tensor, run, tensor + run	$n_s$	0.8	1.2
tensor, tensor + run	$r$	0	1
run, tensor + run	$n_{\text{run}}$	-0.1	0.1

- (III) A running model, where both the spectral index and the running of the spectrum ( $n_{\text{run}}$ ) are allowed to vary.
- (IV) A tensor model, where the spectral index of the scalar perturbations and the tensor-to-scalar amplitude ratio ( $r$ ) are allowed to vary.
- (V) A tensor + running model, where the spectral index, tensor-to-scalar ratio, and running all vary.

The priors on the parameters in these models are given in Table I.

In this analysis we used measurements of the CMB temperature and polarization power spectra from both the WMAP 5 yr [2] and 7 yr [3] releases, to explore how WMAP has improved in its ability to distinguish between different models of the primordial power spectrum. A compilation of WMAP 7 yr and ground-based CMB experiments (Arcminute Cosmology Bolometer Array Receiver [18], Cosmic Background Imager [19], and BOOMERanG [20]), along with the Sloan Digital Sky Survey Data Release 7 [21] measurements of the galaxy clustering power spectrum, was also studied.

We used nested sampling [22] to compute the evidence values and posterior distributions for the different models, making use of COSMONEST [4] and MULTINEST [23] as additional modules for the COSMOMC [24] analysis code.

The evidence values for the various models and different data compilations are given in Table II. We find that the tilted model is the favored model for all data compilations, except for WMAP 7 yr + ext where it is tied with the tensor + running model within the evidence uncertainties. The tilted model is favored over the HZ with strong, but not decisive, evidence (as defined in the Jeffreys’ scale) using the combined data sets. The tensor model is disfavored compared to the HZ using WMAP 5 yr, and 7 yr data alone, but becomes mildly favored when the other data is

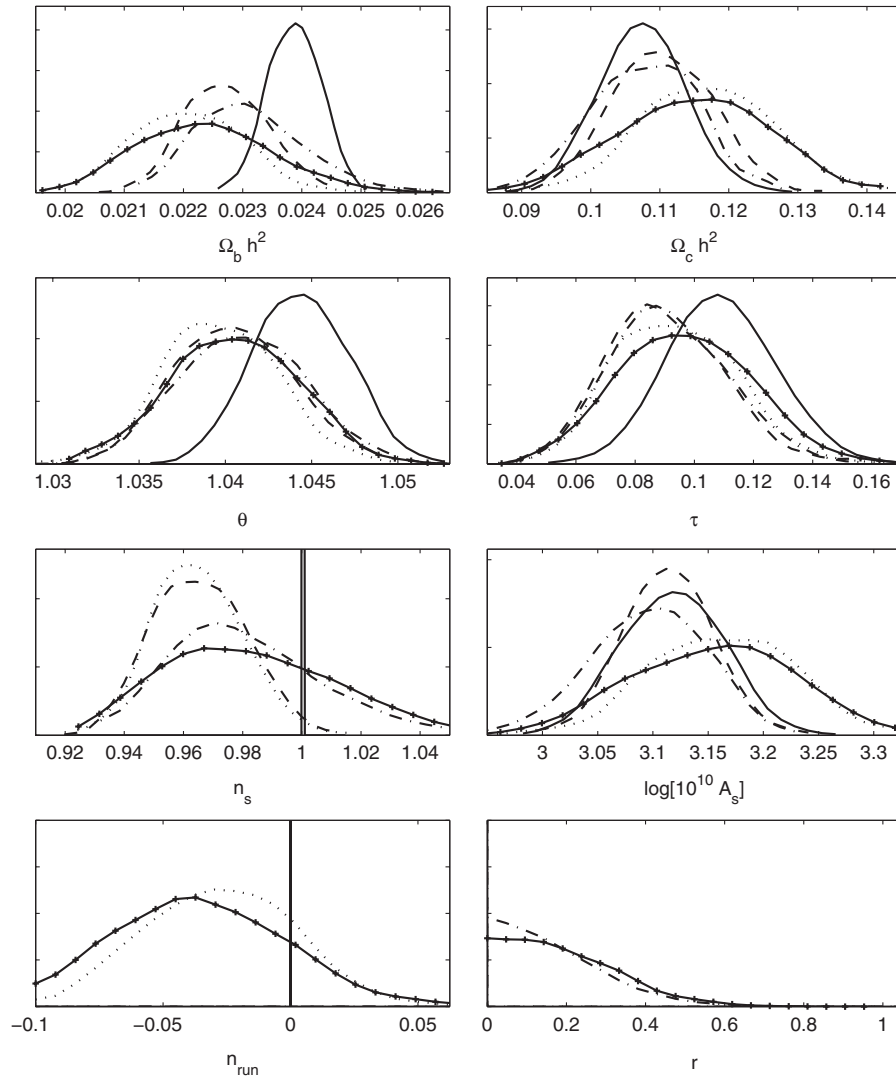


FIG. 1. The posterior probability distributions for the different models, using only the WMAP 5 yr data. The models are HZ (solid lines), tilted (dashed lines), running  $n_s + n_{\text{run}}$  (dotted lines), inflation  $n_s + r$  (dot-dashed lines), and tensor + running (solid lines with crosses).

included (though the evidence differences are too small to be conclusive). The running model has approximately the same evidence as the HZ for WMAP 5 yr and 7 yr, but this increases to be about the same as the tilted model when the extra data sets are added.

TABLE II. (log-)evidence differences for the different models, compared to the HZ model for that data compilation. Positive values mean that the model is favored over HZ. The unnormalized evidence values for the HZ model are  $-1346.3$  (for WMAP 5 yr),  $-3754.5$  (for WMAP 7 yr), and  $-3834.3$  (for WMAP 7 yr + ext).

Model		Datasets		
		WMAP 5 yr	WMAP 7 yr	WMAP 7 yr + ext
I.	HZ	$0.0 \pm 0.1$	$0.0 \pm 0.2$	$0.0 \pm 0.2$
II.	$n_s$	$0.5 \pm 0.2$	$1.0 \pm 0.2$	$3.0 \pm 0.2$
III.	$n_s + n_{\text{run}}$	$-0.1 \pm 0.2$	$0.4 \pm 0.2$	$3.4 \pm 0.2$
IV.	$n_s + r$	$-1.3 \pm 0.1$	$-0.9 \pm 0.2$	$0.8 \pm 0.2$
V.	$n_s + n_{\text{run}} + r$	$-1.1 \pm 0.2$	$-0.7 \pm 0.2$	$2.2 \pm 0.2$

The question of what the probability of  $n_s = 1$  is (i.e., the probability of the HZ model) is one of model selection. Combining the evidence values with model prior probabilities, we can compute the normalized model posterior for each of the models using a normalized version of Eq. (2)

$$P(\mathcal{M}_i|D) = \frac{P(D|\mathcal{M}_i)P(\mathcal{M}_i)}{\sum_k P(D|\mathcal{M}_k)P(\mathcal{M}_k)}. \quad (7)$$

Assuming equal prior probabilities for each of the models,  $P(\text{HZ}) = 0.24$  for WMAP 5 yr,  $0.164$  for WMAP 7 yr, and  $0.016$  for WMAP 7 yr + ext. So even WMAP 7 yr data is not strong enough to exclude an HZ spectrum by itself, but the addition of extra data reduces its model probability by a factor of 10, reducing it to less than 2%.

Other recent papers [8,9] have also computed the Bayes factors for models of the primordial power spectrum, using

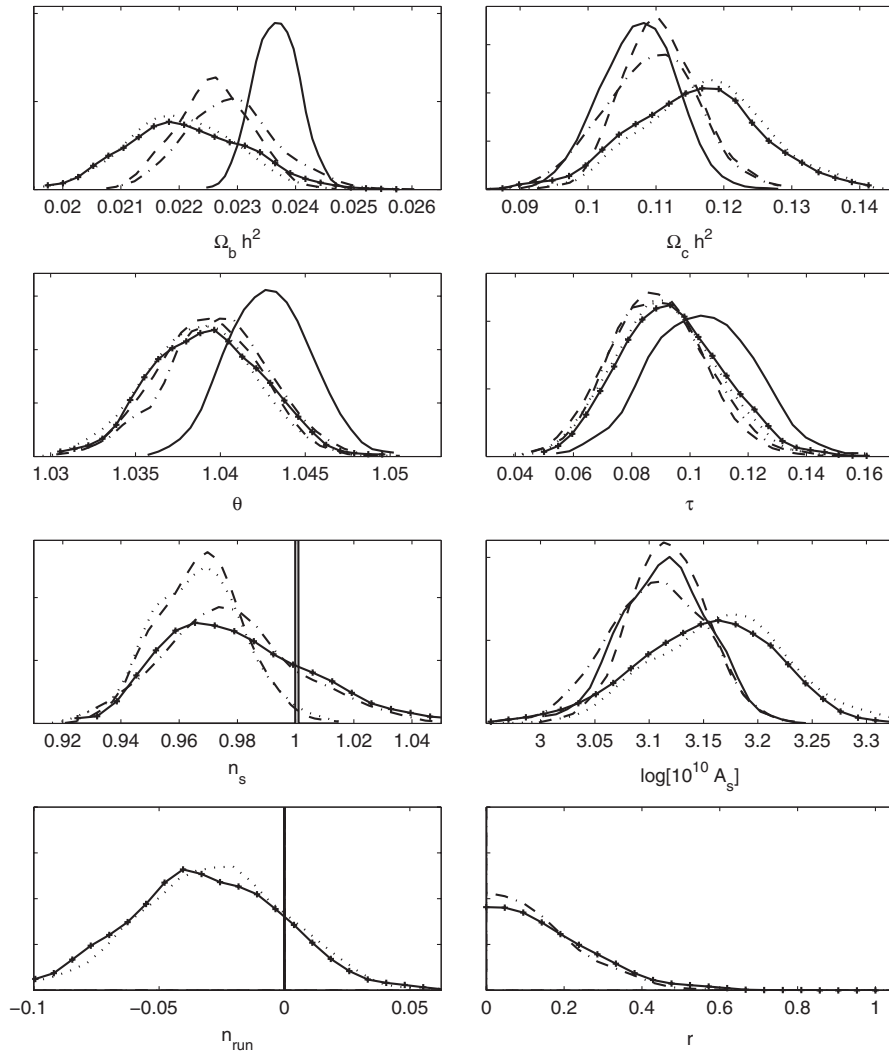


FIG. 2. The posterior probability distributions for the different models, using only the WMAP 7 yr data. The models are HZ (solid lines), tilted (dashed lines), running  $n_s + n_{\text{run}}$  (dotted lines), inflation  $n_s + r$  (dot-dashed lines), and tensor + running (solid with crosses).

the WMAP 5 yr data and other extra data sets. Though it is difficult to do a direct comparison of raw evidence values between our analysis and others (owing to slightly different choices of priors), their basic conclusions are the same as ours: tilted models are favored over HZ, a tensor model is disfavored relative to the others, and running models have roughly the same evidence as tilted models when other data is added in.

The one-dimensional marginalized posteriors for each parameter, model, and data compilation are shown in Figs. 1 (for WMAP 5 yr), 2 (for WMAP 7 yr), and 3 (for WMAP 7 yr + ext). The posteriors are normalized in the usual way, such that the area under each curve is unity. For models where a parameter is not varied (such as the spectral index in the HZ model), a delta function at the appropriate parameter value is the relevant posterior.

In terms of plotting constraints on the power spectrum parameters, the choice of pivot scale  $k_*$  is important. If it is

not optimized, marginalized constraints can appear much weaker than they actually are. We chose the scale that decorrelates the uncertainty on the tilt and the running (in the running model), following the method described in Ref. [25]. This scale is found to be  $0.013 \text{ Mpc}^{-1}$  for the WMAP 5 yr and 7 yr data and  $0.015 \text{ Mpc}^{-1}$  for the WMAP + ext data set.

For the Bayesian model averaging, we assume that the prior model probabilities are equal. Variations of this assumption could readily be explored using the quoted evidence values; for instance one might want to down-weight HZ as it is not based on a physical model, or models with running as inflationary models with large running are hard to construct. Model averages are carried out by combining the posterior samples from different models weighted by the appropriate model probability.

We do not show a model-averaged result for  $A_s$ , as we did not optimize the pivot scale for the amplitude; in the

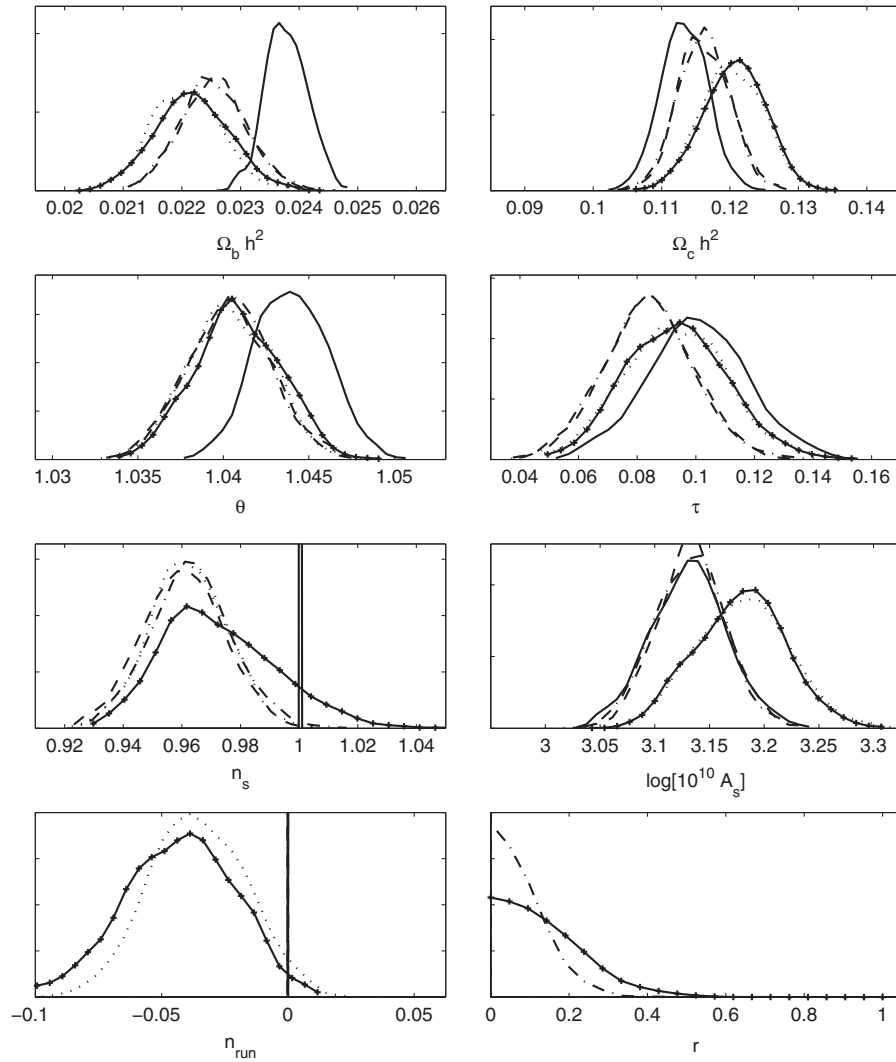


FIG. 3. The posterior probability distributions for the different models, using only the WMAP 7 yr data plus other data sets. The models are HZ (solid lines), tilted (dashed lines), running  $n_s + n_{\text{run}}$  (dotted lines), inflation  $n_s + r$  (dot-dashed lines), and tensor + running (solid with crosses).

figures one sees that the central amplitude is shifted in the running models. This means that the amplitude is best determined at some other scale, and a model averaging should only be carried out at that pivot scale if constraining power is not to be lost. In any case, one can see by eye that model averaging will have little effect on the constraints on  $A_s$ , which is well determined in all models.

The tilt  $n_s$  is more interesting, as it is not so well determined due the residual probability that HZ is correct. Its model-averaged posterior is shown in Fig. 4. Note that in analyses with WMAP data alone the HZ “spike” is prominent in the model-averaged posteriors, containing a significant fraction of the posterior probability. Only once other data are brought in does its effect become small. From the complete data compilation, we find that the model-averaged limits on the tilt at the pivot scale are  $0.940 < n_s < 1.000$  (95% credible interval), the upper

limit being precisely at one as it happens to fall within the delta-function component from the HZ model.

Variations of prior assumptions can modify these results somewhat. Changes to the prior ranges of parameters common to all models, such as  $h$  and  $\tau$ , will have no effect, at least as long as the data constrains the values to lie well within the prior as it does in these cases. Modifying the priors on the parameters that are varied only in some models can shift the results. As an example, we consider doubling the prior range of  $n_s$ , to  $[0.6, 1.4]$ . As the added range fits the data poorly, it has negligible likelihood and this halves the evidence of models in which  $n_s$  varies, i.e., their log evidences in Table II are reduced by  $\ln 2 \sim 0.69$ , which changes the quantitative outcome but not the qualitative one. If one recomputes the 95% confidence range of  $n_s$  under this assumption the range is unchanged. Changes to the assumed prior model probabilities can be handled similarly.



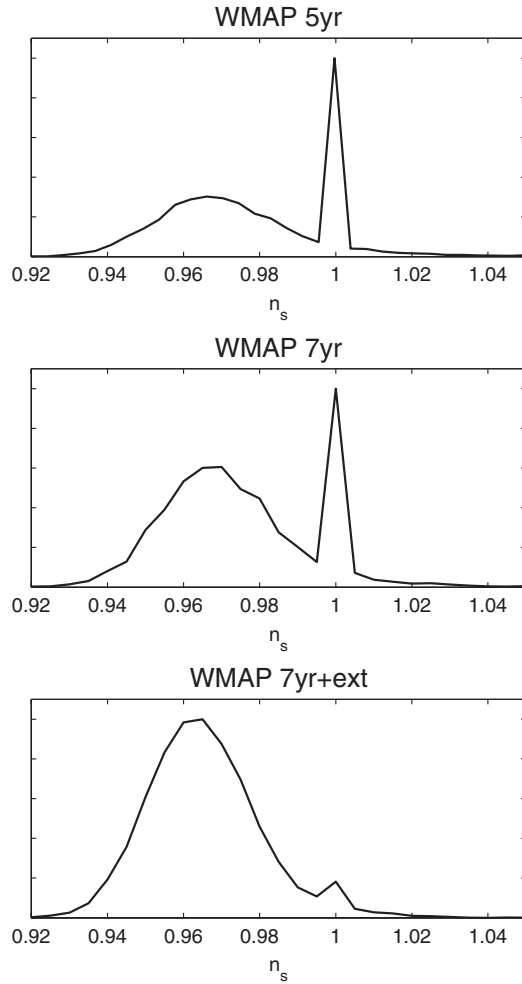


FIG. 4. The model-averaged posterior distributions for the spectral index  $n_s$ , using the WMAP 5 yr data only (top panel), the WMAP 7 yr data only (middle panel), and the WMAP 7 yr + ext compilation (bottom panel). The probability distribution includes a delta function around  $n_s = 1$ , artificially broadened in the plot by the binning process.

Finally, we consider the tensors. As they are not detected, model uncertainty can have a significant impact by lending support to models in which they are entirely absent, i.e.,  $r$  is precisely zero. If we average over all five models, the model-averaged 95% upper limit on the tensor-to-scalar ratio is  $r < 0.16$  (again at the pivot scale). This is indeed somewhat tighter than results from individual models (e.g., the equivalent upper limit from the inflation model is 0.18) because it allows for the possibility of no tensors. Nevertheless, this result is clearly highly prior dependent, and would, for instance, change if one decided that a logarithmic prior on  $r$  were more appropriate.<sup>2</sup>

<sup>2</sup>In practice a logarithmic prior on  $r$  puts almost all the prior model probability at very small  $r$  values, yielding results near identical to a tensorless model provided tensors are not detected.

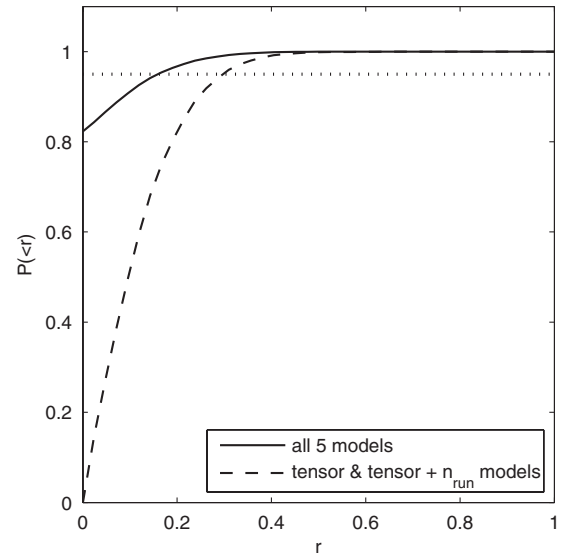


FIG. 5. The model-averaged cumulative probability distribution for the tensor-to-scalar ratio  $r$  using the WMAP 7 yr + ext compilation. The solid curve gives the probability averaged over all five models, whereas the dashed curve gives the probability averaged just over those models where  $r$  is varied (tensor and tensor + running). The dotted line shows the 95% credible limit.

An alternative tensor limit can be obtained by averaging only over the two models which permit tensors, which gives  $r < 0.30$ . The cumulative model-averaged probabilities for  $r$  under both assumptions are shown in Fig. 5. It is clear that any upper limit quoted on the tensor fraction has significant model and prior uncertainty, as well as observational uncertainty.

#### IV. CONCLUSIONS

We have illustrated the methodology of Bayesian model averaging using primordial power spectrum estimation. From a Bayesian viewpoint, the purpose of all data analysis is to start with prior information, perhaps entirely theoretically motivated, and take data of increasing quality until the data likelihood is able to convincingly overcome the prior uncertainty. Bayesian model averaging allows us to include the prior model uncertainty as well as the prior parameter uncertainty in this process, and hence offers a more complete incorporation of theoretical uncertainties.

Currently, data are unable to decisively distinguish amongst different models for the primordial perturbations, with all five that we discuss remaining viable at some level. Despite that, parameters such as the spectral amplitude that are very accurately measured are quite unaffected by model uncertainty. Parameters moderately well determined, such as  $n_s$ , can see significant effects from model uncertainty, while undetermined parameters such as  $r$  are naturally the most sensitive to the increased incorporation of uncertainties.

Bayesian model averaging can be used beyond cosmological models to any problem in cosmology or astrophysics where the underlying model is uncertain. This may be uncertainty as to the nature of the physical object (e.g., unresolved galaxies of different species contributing to a background) or with regards to the data analysis (e.g., supernova light curve analysis, where a number of different light curve fitters are available).

As with any Bayesian analysis, the results will have some dependence on the choices of priors, including the model prior probabilities. We should not be afraid of this; the opportunity to choose appropriate priors is our chance to deploy our physical intuition. Readers who prefer different priors are welcome to recalculate if they wish; this is particularly easy for model priors as the evidence ratios we quoted are all that is needed. We also briefly discussed a

modification of parameter priors; a full analysis should explore the reasonable range of prior possibilities. Only by combining the full range of prior uncertainties, at both parameter and model level, with observational uncertainties, can one obtain the full picture of current understanding.

## ACKNOWLEDGMENTS

The authors were supported by the Science and Technology Facilities Council (Grant No. ST/F002858/1). We thank Mike Hobson, Martin Kunz, and Roberto Trotta for helpful discussions. We acknowledge use of the Archimedes computing cluster at the University of Sussex, supported by funds from SRIF3, and the COSMOS supercomputer in Cambridge, supported by SGI, Intel, HEFCE, and STFC.

- 
- [1] D. N. Spergel *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **170**, 377 (2007); G. Hinshaw *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **170**, 288 (2007); L. Page *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **170**, 335 (2007).
  - [2] J. Dunkley *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **180**, 306 (2009); G. Hinshaw *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **180**, 225 (2009); E. Komatsu *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **180**, 330 (2009); M. R.olta *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **180**, 296 (2009).
  - [3] E. Komatsu *et al.* (WMAP Collaboration), arXiv:1001.4538; D. Larson *et al.* (WMAP Collaboration), arXiv:1001.4635.
  - [4] P. Mukherjee, D. Parkinson, and A. R. Liddle, *Astrophys. J. Lett.* **638**, L51 (2006).
  - [5] D. Parkinson, P. Mukherjee, and A. R. Liddle, *Phys. Rev. D* **73**, 123523 (2006).
  - [6] R. Trotta, *Mon. Not. R. Astron. Soc.* **378**, 72 (2007).
  - [7] M. Beltrán, J. García-Bellido, J. Lesgourgues, A. R. Liddle, and A. Slosar, *Phys. Rev. D* **71**, 063532 (2005); M. Bridges, A. N. Lasenby, and M. P. Hobson, *Mon. Not. R. Astron. Soc.* **369**, 1123 (2006); M. Kunz, R. Trotta, and D. Parkinson, *Phys. Rev. D* **74**, 023503 (2006); M. Bridges, A. N. Lasenby, and M. P. Hobson, *Mon. Not. R. Astron. Soc.* **381**, 68 (2007); R. Trotta, *Mon. Not. R. Astron. Soc.* **375**, L26 (2007); M. Bridges, F. Feroz, M. P. Hobson, and A. N. Lasenby, *Mon. Not. R. Astron. Soc.* **400**, 1075 (2009); I. Sollom, A. Challinor, and M. P. Hobson, *Phys. Rev. D* **79**, 123521 (2009); J. Valiviita and T. Giannantonio, *Phys. Rev. D* **80**, 123516 (2009).
  - [8] M. Kawasaki and T. Sekiguchi, *J. Cosmol. Astropart. Phys.* **02** (2010) 013.
  - [9] M. Kilbinger *et al.*, *Mon. Not. R. Astron. Soc.* **405**, 2381 (2010).
  - [10] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, *Stat. Sci.* **14**, 382 (1999).
  - [11] R. E. Kass and A. E. Raftery, *J. Am. Stat. Assoc.* **90**, 773 (1995).
  - [12] H. Jeffreys, *Theory of Probability* (Oxford University Press, Oxford, 1961), 3rd ed..
  - [13] A. Jaffe, *Astrophys. J.* **471**, 24 (1996); P. S. Drell, T. J. Loredo, and I. Wasserman, *Astrophys. J.* **530**, 593 (2000); M. V. John and J. V. Narlikar, *Phys. Rev. D* **65**, 043506 (2002); A. Slosar *et al.*, *Mon. Not. R. Astron. Soc.* **341**, L29 (2003); T. D. Saini, J. Weller, and S. L. Bridle, *Mon. Not. R. Astron. Soc.* **348**, 603 (2004); P. J. Marshall, M. P. Hobson, and A. Slosar, *Mon. Not. R. Astron. Soc.* **346**, 489 (2003); A. Niarchou, A. H. Jaffe, and L. Pogosian, *Phys. Rev. D* **69**, 063515 (2004); B. A. Bassett, P. S. Corasaniti, and M. Kunz, *Astrophys. J. Lett.* **617**, L1 (2004); P. Mukherjee, D. Parkinson, P. S. Corasaniti, A. R. Liddle, and M. Kunz, *Mon. Not. R. Astron. Soc.* **369**, 1725 (2006).
  - [14] G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces* (Wiley, New York, 1987), p. 424.
  - [15] P. J. Marshall, M. P. Hobson, and A. Slosar, *Mon. Not. R. Astron. Soc.* **346**, 489 (2003); A. R. Liddle, P. Mukherjee, D. Parkinson, and Y. Wang, *Phys. Rev. D* **74**, 123506 (2006).
  - [16] W. H. Jefferys, Eichhorn Memorial Lecture 2000, available at <http://quasar.as.utexas.edu/Papers.html>; T. G. Barnes, W. H. Jefferys, J. O. Berger, P. J. Mueller, K. Orr, and R. Rodriguez, *Astrophys. J.* **592**, 539 (2003); **611**, 621(E) (2004); S.-K. Min and A. Hense, *Geophys. Res. Lett.* **33**, L08708 (2006); J. Debosscher, L. M. Sarro, C. Aerts, J. Cuypers, B. Vandenbussche, R. Garrido, and E. Solano, *Astron. Astrophys.* **475**, 1159 (2007).
  - [17] R. Harrison, *Phys. Rev. D* **1**, 2726 (1970); Ya. B. Zel'dovich, *Mon. Not. R. Astron. Soc.* **160**, 1 (1972).
  - [18] C. L. Reichardt *et al.*, *Astrophys. J.* **694**, 1200 (2009).
  - [19] J. L. Sievers *et al.*, *Astrophys. J.* **660**, 976 (2007).
  - [20] W. C. Jones *et al.*, *Astrophys. J.* **647**, 823 (2006).

- [21] B.A. Reid *et al.*, *Mon. Not. R. Astron. Soc.* **404**, 60 (2010).
- [22] J. Skilling, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by R. Fischer *et al.*, AIP Conf. Proc. No. 735 (AIP, New York, 2004), p. 395, available at <http://www.inference.phy.cam.ac.uk/bayesys/>.
- [23] F. Feroz and M. P. Hobson, *Mon. Not. R. Astron. Soc.* **384**, 449 (2008); F. Feroz, M. P. Hobson, and M. Bridges, *Mon. Not. R. Astron. Soc.* **398**, 1601 (2009).
- [24] A. Lewis and S. Bridle, *Phys. Rev. D* **66**, 103511 (2002), code available from <http://cosmologist.info/cosmomc>.
- [25] M. Cortes, A. R. Liddle, and P. Mukherjee, *Phys. Rev. D* **75**, 083520 (2007).